DOCUMENT RESUME

ED 328 071 FL 019 036

AUTHOR Schils, Erik; Weltens, Bert
TITLE Modelling and Assessing Foreign Language Loss.
PUB DATE 91
NOTE 13p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Cloze Procedure; Foreign Countries; *French; *Language Maintenance; *Language Proficiency; Language Research; *Language Skill Attrition; *Language Tests; Linguistic Theory; Models; Multiple Choice Tests; Secondary Education; Second Language Learning; *Test Format; Testing; Vocabulary
IDENTIFIERS *Netherlands

ABSTRACT

A study investigated the retention and loss of school-learned foreign language skills. Subjects were 150 students of Dutch secondary schools with 4 or 6 years of French language instruction and 0, 2, and 4 years of disuse. Nine tests of general language proficiency, listening and reading comprehension, phonology, vocabulary, and grammar were administered. Analysis of the results indicates that: (1) some integrative tests may be more sensitive than others to changes in universal language proficiency; (2) the multiple choice cloze test used here was a more valid indicator of French proficiency than the listening and reading comprehension tests; (3) the cloze test failed to reveal loss of vocabulary or grammar; and (4) contrasts in the design-dependent and design-independent analyses imply a need to separate validity and reliability studies of language tests from their application in research or education. A brief list of references is appended. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED328071

MODELLING AND ASSESSING FOREIGN LANGUAGE LOSS

Erik Schils & Bert Weltens

University of Nijmegen

The Netherlands

Accepted for Publication in

Ludo T. Verhoeven & John H.A.L. de Jong

The Construct of Language Proficiency:

Applications of Psychological Models to Language Assessment

Amsterdam: John Benjamins

1991

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
B. Weltens

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

FL019036

Modelling and assessing foreign language loss

Erik Schils & Bert Weltens

The debate on the construct of language proficiency has been characterized by a pendular movement going from models incorporating a large number of factors, to the other extreme of Oller's unitary competence hypothesis (for an overview, see, e.g., De Jong, 1990; Masters, 1990). Most of the contributions to this debate have been based on either synchronic research or diachronic language acquisition research. In this chapter we discuss the construct of language proficiency in relation to a diachronic study of language attrition. More specifically we will be dealing with foreign language loss, that is, the loss of foreign language proficiency after a period of disuse.

The study of foreign language loss represents a fairly new line of research in applied linguistics (Clark; 1982, Lambert and Kroon, 1987; Lambert and Moore, 1986; Weltens and Cohen, 1989). Clark (1982: 140) argues that a "comprehensive and maximally useful" study of language loss "would need to incorporate both general proficiency testing and diagnostic feedback testing, for quite distinct but complementary purposes" (p. 140). An indication of general proficiency is necessary to obtain insight into the significance of the loss in terms of everyday use; whereas, diagnostic tests are required to obtain information on "the particular features involved in language performance at a given level of general proficiency". As Moorcroft and Gardner (1987: 328) point out, it is the latter type of information that might enable "educators and researchers (...) to develop a better understanding of processes operating in the acquisition of language that they might strengthen to promote second-language retention".

In a recent research project, where both integrative and discrete point tests were used, it was found that little or no loss of school-learned French could be detected after periods of disuse as long as four years (Weltens, 1989; Weltens, Van Els, and Schils, 1989). The only skills that appeared to have clearly suffered from disuse were grammatical skills. Overall (receptive) proficiency, as measured by a MC cloze test, remained virtually unchanged across the interval investigated. Phonology tests, and tests for listening and reading comprehension even exhibited significant *increases* in proficiency over time. Two vocabulary tests produced opposite results, with an aural vocabulary test indicating some gain, and a written vocabulary test indicating some loss.

In addition to the tests listed above, a lexical-decision experiment was carried out which showed that, as far as the *speed* of lexical access and retrieval is concerned, an interval of two years already leads to a significant increase in reaction times. In other words, the results suggested that it is not so much the success of vocabulary retrieval, but the speed of that process which is affected by disuse. (For a more detailed account of the experiment and its implications for the other results, see Weltens, 1989 :94-96; Weltens & Grendel, 1989.)

Two aspects of the results of this research seemed irreconcilable:

1. some loss was detected in vocabulary and, in particular, in grammar, but no loss was observed in general proficiency as measured by a multiple choice cloze test;
2. no change was detected in general proficiency as measured by a multiple choice cloze test, versus an observed increase in general proficiency as measured by multiple choice listening and reading comprehension tests.

In order to throw some light on these apparently contradictory results some further analyses using the same data set are presented in this chapter. The data used in the analyses presented here originate from a research project on the retention and loss of school-learned French foreign-language skills. We will limit our description of the project to an outline of the relevant aspects; for a detailed account of the project and its results, the reader is referred to Weltens (1989), and Weltens, Van Els, and Schils (1989).

Method

Subjects

The project involved 150 students and former students of Dutch secondary schools. The students were evenly distributed across a 2 x 3 factorial design. The two factors in the design were education in French (6 or 4 years) and period of disuse since the end of French training (0, 2, 4 years). The design is presented in table 1.

Table 1
Design of the investigation

No. of years disuse		0	2	4
No. of years of	6	A	B	C
French education	4	D	E	F

The subjects, 25 in each cell, were selected in such a way that they did not differ greatly in terms of learning career, general language ability, number of foreign languages -- other than French -- studied in the higher grades of secondary school, number of years of training in Latin, and post-course exposure to French.

Tests

The following nine tests were administered:

- a multiple choice cloze test (82 items; α = .90);
- two integrative multiple choice tests, one for listening and one for reading comprehenion (α = .79 and α = .82 respectively);
- three aural and three written tests, testing for phonology, vocabulary, and grammar (with α ranging from .69 to .80).

All tests were constructed in the course of the project, except the integrative listening and reading tests, which had been developed by CITO, the Dutch National Institute for Educational Measurement, as part of the national examinations for French.

The multiple choice cloze test consisted of two texts from which every fifth word had been deleted, resulting in a total of 82 blanks. For each blank three options were given: the word from the original text, and two distracters selected from the responses to the open-ended version of the test given by a group of subjects that was comparable to the subject sample in the actual investigation.

The listening comprehension (LC) test consisted of 58 text fragments presented on audio tape. In the short pause following each fragment the

subjects had to answer a multiple choice question about the fragment. Multiple choice questions were presented in print, and they were also in French.

The reading comprehension (RC) test consisted of five text fragments taken from French newspapers with a total of 50 accompanying multiple choice questions. As with the listening comprehension test, all test materials were in French.

The phonology tests consisted of 34 items each. In the aural version, each item consisted of a triad, namely a minimal pair, one member of which was presented twice. To give a few examples:

pas - bas - bas
ton - temps - ton
bige - bige - biche

The subjects were asked to tick on the answer sheet which of the three words sounded identical. Of the total of 40 items, 34 were actual target items; the other six were filler items that consisted of three identical words (e.g., *blanche - blanche - blanche*).

The written phonology test was a 'same/different' test consisting of 68 pairs of words, 34 target items and 34 filler items. The target items tested the same phonemic oppositions as the aural version and required, therefore, 'different' responses. The filler items consisted of pairs of homophones -- with the result that all pairs differed in spelling -- and they required a 'same' response. Examples of the filler items are the following:

pot - peau
en - an
hors - or

The vocabulary tests consisted of 40 items each. A French sentence was presented, plus its translation in Dutch, but in the Dutch rendering the target word was left out. For example:

Marianne a mal à la cheville.
Marianne heeft pijn aan haar ..... .
[Marianne's ..... is hurting.]

In the aural vocabulary test the French sentence was presented on audio tape and the Dutch rendering on the answer sheet; in the written test both were given in print.

The grammar tests consisted of 40 items each. In the aural version a sentence was presented on audio tape twice, with the target replaced by a beep; the answer sheet contained three alternatives to choose from. In the written version both sentence and answering alternatives were given on paper. An example item is the following:

Voulez-vous ..... votre valise à la gare?
0 laisser
0 laissé
0 laissez

Procedure.
The subjects were tested in a language lab on two consecutive days. Two sessions were considered necessary because the total administration time, including breaks, amounted to seven hours. The aural tests were played to the subjects via headphones. Before the actual tests were administered, the

subjects took a short listening test as a warm-up. This test tested for the use of place indications such as *en face de*, *loin de*, and *tout droit*, which did not occur as targets in any of the actual tests. The warm-up session was included because all subjects had hardly used French outside the school, and we wanted to avoid them having a 'cold start'.

Analyses

First of all, a factor analysis was carried out on all test scores for groups A, C, D, and F from table 1, that is, the end-points of our design. We included only these end-points because our data indicated that the influence of the education factor outweighed by far the influence of the time factor. By including the end-points we allowed for a maximum influence of the time factor.

The reliability of the results was assessed by means of the so-called *bootstrap* approach (Chatterjee, 1984; Schils, 1988). The bootstrap is a computer-based method for the study of the sampling distribution of statistics, especially in situations where an analytical sampling theory of the particular statistic is (as yet) lacking, or where the assumptions underlying an existing sampling theory are not met. It yields an approximation of the sampling distribution and in particular the sampling variability of the statistic at hand by calculating this statistic for a large number of what might be called "autogenous reproductions" of the empirical sample. This term refers to a series of random samples drawn, with replacement, from a finite population which exactly mirrors the empirical sample. The bootstrap estimate of the standard error of the statistic, then, is the standard deviation of all these "reproduced" values; it can further be used in performing statistical tests and in constructing confidence intervals. Unfortunately, the scores for listening and reading comprehension could not be used in this analysis, because these two tests had not been administered to group D.

The second analysis we carried out was a multiple regression analysis in which we tried to answer the question of why the cloze test indicated no change in general proficiency, whereas the tests for listening and reading comprehension indicated an increase in proficiency. On earlier occasions, it was suggested that the increase might be due to an increase in "universal" language proficiency -- not just French -- as a result of continued learning of other foreign languages and continued acquisition, to some degree, of the mother tongue (cf. Weltens, 1989). If this suggestion is correct, it still remains unclear why the effect of this universal language proficiency would be so much stronger on two out of three integrative tests. The least we would have to demonstrate is that the cloze test is relatively insensitive to differences in universal language proficiency.

Results

On the basis of the criterion of number of eigenvalues greater than 1, the principal component analysis extracted two factors from the data. After varimax rotation, the factor loadings shown in table 2 emerged. The bootstrap estimates of the corresponding standard errors (based on 100 re-samplings) are given between brackets.

Table 2
Results of a factor analysis of all test scores:
factor loadings (bootstrap standard errors)

	Factor 1	Factor 2
Grammar (aural)	.85 (.028)	.22 (.085)
Cloze	.83 (.039)	.27 (.076)
Grammar (written)	.83 (.044)	-.01 (.099)
Vocabulary (written)	.82 (.039)	-.01 (.114)
Vocabulary (aural)	.68 (.076)	.50 (.122)
Phonology (aural)	.12 (.068)	.86 (.033)
Phonology (written)	.07 (.072)	.85 (.033)

The total of 100 bootstrap replications of the analysis indicated great stability in the number of factors extracted. No less than 99 re-samplings yielded a two-factor solution. In the one remaining case, a three-factor solution was obtained in which the first factor of table 2 was broken down into grammar on the one hand, and vocabulary plus cloze on the other. The partition of the tests across the factors was very stable as well, with the exception of the aural vocabulary test: among the 99 two-factor solutions, the partition shown in table 2 emerged 79 times; in 19 cases aural vocabulary joined the "phonological" cluster. This is, of course, in line with the small difference in the corresponding factor loadings for this test, as shown in table 2: .68 on factor 1, and .50 on factor 2, together with their relatively large bootstrap standard errors (.076 and .122 respectively). The remaining solution was an uninterpretable one, with grammar and written vocabulary as one factor, and the rest lumped together in the second.

In interpreting this factor pattern one should be aware of the fact that it does not only reflect the interrelatedness of the tests per se, but that it is also partly determined by the sensitivity of the tests to the design factors education and disuse. This entails some equivocality as to the correct interpretation. From a substantive point of view, for instance, the interpretation of factor 2 seems quite straightforward, as it corresponds to the cluster of phonology tests. The aural vocabulary test, however, although its adhesion to the first factor seems to dominate, has a respectable loading on factor 2 as well. Although, from a substantive point of view, this could be interpreted as the degree to which the aural vocabulary test calls upon phonological skills, it should also be noted that, in terms of "susceptibility to the design," it shares with the phonology tests the tendency to increase under disuse.

An even greater degree of equivocality is attached to the interpretation of the first factor. From a substantive point of view, one could on the one hand argue in favour of labelling it "general proficiency in French", whereas on the other hand a label such as "vocabulary plus grammar" could be chosen, as cloze tests have been repeatedly shown to be tests of, primarily, those two aspects of language proficiency (see, e.g., Jonz, 1988). With respect to the design aspects, the "factor 1 tests" exceeded the "factor 2 tests" in sensitivity to education, i.e., they exhibited much larger main effects for education (cf. Weltens, 1989).

Although the LC and RC tests could not be entered into the factor analysis, we could of course compute the correlations between the scores for those tests and the factor scores that had emerged from the factor analysis. These correlations, together with the bootstrap estimates of their standard errors which emerged from 100 replications, are given in table 3.

Table 3
Pearson correlations between LC and RC
and the factor scores (standard errors)

	Factor 1	Factor 2
Listening comprehension	.69 (.047)	.20 (.108)
Reading comprehension	.70 (.065)	.21 (.090)

The results clearly show that both tests were related to factor 1, and only very weakly to factor 2. From the substantive point of view, this finding is not surprising, given the fact that these tests are also known, just like the cloze test, to rely on lexical and grammatical skills to some degree. RC and LC did show resistance against disuse: RC improved significantly for both education levels; LC improved slightly for the subjects with 6 years of French education, and significantly so for those with 4 years of French education. Still, according to the design-related interpretation one would have to conclude that for these tests, "susceptibility to education" seems to predominate, with the result that they link up with factor 1.

In order to disambiguate the above-mentioned results, the data were re-analyzed after partialling out the influence of the design factors. The test scores were standardized groupwise prior to calculating their intercorrelations. The resulting "pooled within-groups correlation matrix" was then submitted to factor analysis again, along with a bootstrap evaluation of the results, again based on 100 re-samplings. As the input correlation matrix in a sense could be characterized as a matrix of residual correlations, there is an even greater need for such a reliability check than in the previous analysis.

Table 4
Results of a design-independent factor analysis
of all test scores: factor loadings

	Factor 1	Factor 2	Factor 3
Vocab. (written)	.87	.01	-.09
Vocab. (aural)	.76	.00	.45
Cloze	.67	.46	.09
Grammar (written)	.04	.93	.05
Grammar (aural)	.15	.80	.36
Phon. (aural)	.07	.13	.82
Phon. (written)	.06	.16	.78

The factor analysis of the design-independent correlations resulted -- after varimax rotation -- in the three-factor pattern given in table 4.

Inspection of the partition of tests across factors suggests the labels vocabulary (1); grammar (2), and phonology (3), with cloze loading on vocabulary and, to a lesser degree, also on grammar. As for the third factor, it should also be noted that the aural versions of the tests all have higher loadings on this factor than their written counterparts, which underscores its sound-related character.

The bootstrap evaluation of these results indicated much larger instability than was the case with the original analysis. Therefore, bootstrap estimates of the standard errors of the factor loadings are not presented, as they would lump together figures corresponding to fundamentally different solutions. The instability concerned the following aspects:

1. *The number of factors extracted*: a three-factor solution was obtained in only 57 out of 100 bootstrap replications; the remaining 43 were two-factor solutions.
2. *The partition of tests across factors*: out of the 57 three-factor solutions, the empirically obtained solution of table 4 was replicated in 41 cases; in 7 other cases it was comparable, except that cloze joined the grammatical factor. Among the 43 bi-factorial solutions there were 26 with cloze and the lexical tests on the one hand, and a combination of both phonology and both grammar tests on the other. This latter cluster may at first sight seem a bit odd, but it may well be that it reflects a test-form artefact, in the sense that all these tests were of the multiple choice type. There was a wide range of other solutions, one of which is definitely worth noting here: the design-dependent partition as obtained in the original analysis (cf. table 2) occurred only once in 100 replications.

In an attempt to separate the constant and variable aspects of the bootstrap replications, we counted the number of times (out of 100 bootstrap resamplings) that a pair of tests could be labelled "homofactorial", i.e., the number of times that they both had a loading of at least .40 on one of the extracted factors. The resulting matrix is presented in table 5.

Table 5 "Homofactoriality" of the tests

	Phon-A	Phon-W	Gram-A	Gram-W	Vocab-A	Vocab-W
Cloze	21	24	55	54	*97*	*96*
Phon-A		*96*	55	40	48	8
Phon-W			52	37	49	10
Gram-A				*100*	21	5
Gram-W					6	3
Vocab-A						*100*

If we would, quite arbitrarily, regard a percentage of 95 over repeated bootstrap analyses as a stability requirement for "togetherness", then -- given the power of the present investigation -- we are inclined to say that there is stable evidence for the presence of three factors: phonology, grammar, and vocabulary, with the cloze test clearly joining the lexical cluster. The relationship between cloze and grammar is, if significant at all -- again within the limitations of the power of the present investigation -- of comparable strength as the relationship between the aural tests and

8

phonology, and the possibly test-form-determined relationship (see above) between grammar and phonology.

In conclusion, we might say that the second analysis, the design-independent one, revealed the substantive interpretations of the original analysis in terms of the structure of language proficiency -- i.e., the theme of the present volume -- as design-dependent results. On the other hand, a comparison of the two analyses gives an impression as to how various aspects of language proficiency behave differentially under processes of learning and forgetting.

In view of the results discussed in the previous section, it was an obvious step to investigate to what degree these three tests, in their development with respect to learning and forgetting, make different appeals to the three underlying factors we identified, and -- in view of the above -- also to general language proficiency. As far as the three factors are concerned, the previous analysis had provided us with strong arguments for combining the aural and written versions into one score per factor, with the additional advantage of reducing the number of predictors to be entered into the analysis. As far as universal language proficiency is concerned, we happened to have a measure that we could use as an indicator of it for our subjects, namely in the form of a self-reported average school mark for all modern languages taken together. This value, to which we will be referring as "grade point average for languages" (GPA-L), was entered into the multiple regression analysis as a predictor variable as well. As explained above, the inclusion of the LC and RC tests implied that we could only use the data for 75 of the subjects, viz. groups A, C, and F (see table 1). The results of this analysis are summarized in table 6; it also lists the bootstrap standard errors resulting from 100 replications of the analysis.

Table 6 Beta weights and multiple correlations from a multiple regression analysis (bootstrap standard errors)

	Cloze	LC	RC
Phonology	.07 (.062)	.11 (.075)	.12 (.078)
Lexicon	.53 (.095)	.35 (.095)	.41 (.103)
Grammar	.31 (.077)	.31 (.093)	.20 (.090)
GPA-L	.01 (.066)	.23 (.073)	.29 (.082)
R	.79 (.055)	.75 (.039)	.75 (.046)
R^2	.62	.56	.57

When we compare the beta weights with their bootstrap standard errors, it turns out that the contribution of lexicon and grammar is significant for all three criterion tests, and that GPA-L carries significant weight only in the case of LC and RC. Thus, the data in table 6 suggest that, in the present context, all three integrative tests rely heavily on vocabulary and grammar. Secondly, the data also seem to indicate a greater predictive power of lexicon than of grammar in the case of both cloze and RC. Thirdly, the table also suggests what the essential difference is between cloze on the one hand, and LC and RC on the other: cloze is the only one of the three where GPA-L has no additional predictive power beyond vocabulary and grammar.

In order to assess the stability of these results, and in particular the differences between the relevant beta weights, we further inspected the results of the 100 bootstrap re-samplings, and counted the number of times that the differences were in the same direction as in table 6. The results confirmed the earlier observations: in predicting the cloze test and the RC, there was indeed a tendency for the contribution of lexicon to be stronger than that of grammar (85 and 81 out of 100 respectively); no such difference was found for the LC test. As far as GPA-L is concerned, it turned out to be a stronger predictor of LC and RC than of cloze in all 100 replications. Moreover, there was a tendency for GPA-L to be stronger in predicting RC as compared to LC (72 out of 100); an investigation with substantially more power than the present one would be necessary to decide whether the latter effect is real.

Discussion

The analyses that were carried out did indeed clarify some of the problematic aspects in our data. They showed that one must be careful in interpreting the results of integrative tests when investigating changes in language proficiency, as at least some types of test may be more sensitive to changes in universal language proficiency -- i.e., not just proficiency in the target language in question -- than others. This ties in with the warning raised by Jaspaert and Kroon (1987), who argue that metalinguistic skills and testwiseness are imminent dangers for especially such tests. Our analyses have shown that, at least in this respect, the cloze test is a much more valid indicator of French language proficiency than the LC and RC tests. Strictly speaking we should say this cloze test, since it has been repeatedly pointed out that cloze tests tend to vary both in reliability and validity (see, e.g., Klein-Braley, 1985).

CITO's reading comprehension test -- which contributes no less than 50% to the final mark for each of the modern foreign languages in Dutch secondary schools -- appears to be particularly sensitive to differences in universal language proficiency, while the listening comprehension test tends to hold an intermediate position in this respect. For the RC test, CITO had already received indications that the test might be less suitable for testing adults. After all, it is validated for adolescents taking secondary school exams, and Maas-de Brouwer's (1983) observation that, generally speaking, "adults score significantly higher than younger candidates" (p. 183; our translation) need not really surprise us, therefore. As far as the LC test is concerned, De Jong (1983: 33) has clearly shown, at least for a pretest version of the listening comprehension test, that two thirds of the items in the test do indeed measure listening comprehension, but that there are strong indications that the other third of the test seems more suitable for testing an ability which might be general intelligence or knowledge of the world. As far as the nature of the skill underlying this "other third" is concerned, however, our analyses suggest that it is related to proficiency in other languages.

On the other hand, we are still left with an inconsistency. Although the analyses reported here have clearly shown that our cloze test relies heavily on knowledge of vocabulary and grammar -- the two aspects which have been shown to suffer from disuse -- the cloze test itself failed to reveal any loss. Similar results were obtained by Gardner et al. (1987: 39): they reported "only slight evidence for language attrition in the objective assessments of proficiency and even some apparent suggestion of growth". Therefore, we cannot but agree with their conclusion that "rather than indicating little language loss or even growth, it is probably more likely

that these results underscore the difficulty of measuring change using objective measures".

Finally, a methodological note is in order: the contrast between the design-dependent and the design-independent analyses implies a strong admonition to separate validity and reliability studies of language tests from their application in research or education.

Acknowledgement

The project that the data were derived from was funded by the Dutch Linguistics Foundation, which is sponsored by the Netherlands Organization for Scientific Research (NWO), under grant no. 300-167-006, during the years 1984-1988.

References

Chatterjee, S. (1984). Variance estimation in factor analysis: An application of the bootstrap. British Journal of Mathematical and Statistical Psycholoy, 37, 252-262.

Clark, J. (1982). Measurement considerations in language attrition research. In: R. Lambert & B. Freed (Eds), The loss of language skills. Rowley, MA: Newbury House.

De Jong, J.H.A.L. (1983). Focusing in on a latent trait: An attempt at construct validation by means of the Rasch model. In: J. van Weeren (Ed), Practice and problems in language testing 5: Non-classical test-theory, final examinations in secondary schools. Arnhem: Cito.

De Jong, J.H.A.L. (1990). Response to Masters: Linguistic theory and psychometric models. In: J.H.A.L. de Jong & D.K. Stevenson (Eds), Individualizing the assessment of language abilities. Clevedon: Multilingual Matters.

Gardner, R., Lalonde, R., Moorcroft, R., & F. Evers (1987). Second language attrition: The role of motivation and use. Journal of Language and Social Psychology 6, 29-47.

Hughes, A. (1982). The structure of language proficiency. In: D. Crystal (Ed.), Linguistic controversies. Essays in linguistic theory and practice in honour of F.R. Palmer. London: Edward Arnold.

Jaspaert, K. & S. Kroon (1987). The relationship between global language proficiency tests and language loss. In: F. Beukema & P. Coopmans (Eds), Linguistics in the Netherlands 1987. Dordrecht: Foris Publications.

Jonz, J. (1988). Constraints on the cloze response. Paper presented at the 22nd Annual TESOL Convention, Chicago, March 7-12.

Klein-Braley, C. (1985). Reduced redundancy as an approach to language testing. In: C. Klein-Braley & U. Raatz (Eds), Fremdsprachen und Hochschule. Bochum: Ruhruniversität (AKS-Rundbrief 13/14), 1-13.

Lambert, R. & S. Moore (1986). Problem areas in the study of language attrition. In: B. Weltens et al. (Eds), Language attrition in progress. Dordrecht: Foris Publications.

Maas-de Brouwer, T. (1983). Normeren van centrale examens [Providing norms for national examination papers]. In: W. Weeda (Ed.), Examens in discussie. Groningen: Wolters-Noordhoff.

Masters, G. (1990). Psychometric aspects of individual assessment. In: J.H.A.L. de Jong & D.K. Stevenson (Eds), Individualizing the asssessment of language abilities. Clevedon: Multilingual Matters.

Moorcroft, R. & R. Gardner (1987). Linguistic factors in second language loss. Language Learning 27, 327-340.

Oller, J. W., Jr. (1979). Language tests at school. London: Longman.

Schils, E. (1988). Bootstraps, een nieuw statistisch alternatief [Bootstraps, a new statistical alternative]. Toegepaste Taalwetenschap in Artikelen 31, 45-56.

Spolsky, B. (1990). Social aspects of individual assessment. In: J.H.A.L. de Jong & D.K. Stevenson (eds), Individualizing the asssessment of language abilities. Clevedon: Multilingual Matters.

Weltens, B. (1989). The attrition of French as a foreign language. (Also available: Dordrecht Foris Publications) Ph. D. Dissertation, University of Nijmegen.

Weltens, B. & A. Cohen (1989). Language attrition research: An introduction. Studies in Second Language Acquisition 11, 127-133.

Weltens, B. & M. Grendel (1989). Taalverlies en woordherkenning in het Frans [Language loss and word recognition in French]. Toegepaste Taalwetenschap in Artikelen 34, 76-80.

Weltens, B., T. van Els & E. Schils (1989). The long-term retention of French by Dutch students. Studies in Second Language Acquisition 11, 205-216.